Independent Peer Review Report on the combined status of blue and deacon rockfishes and California scorpionfish

By

Robin Cook

Prepared for
Center for Independent Experts

Contents

Executive Summary

i. Assessments of blue/deacon rockfish (BDR) in Oregon and California, and California scorpionfish were reviewed during a formal, public, meeting of fishery stock assessment experts from 24th-28th July 2017. Two CIE reviewers were included in the review panel.

ii. The assessment of BDR and scorpionfish represent the best science available given the existing data. The limited amount of age data and lack of informative abundance indices means that it is difficult to have high confidence in the estimated stock status. Sensitivity analyses indicate that alternative explanations of the data are possible which change perceived status. The problem is most severe in the case of California BDR, where the estimated stock recovery is highly uncertain.

iii. The approach to estimating CPUE abundance indices from fishery data was thorough and appears to be the best available. Disappointingly, the resulting indices did not appear to contain much population signal and tended not to contribute much to the estimated stock biomass trajectory. In the Oregon BDR assessment removing all the surveys from the assessment made no difference to the estimated biomass. If possible, fishery independent surveys should be developed to calibrate estimates of biomass in the most recent year.

iv. The catch data are influential in the assessment but are treated as exact and fixed in the model. While this is probably a necessary assumption, it is clearly unrealistic. A demanding sensitivity analysis is required where plausible alternative catch streams are generated stochastically and used to test the model.

v. Priority should be given to the collection and processing of more age samples from all three species. This needs to be maintained to create a coherent time series of observations.

vi. Natural mortality is the largest component of total mortality in these stocks and will drive much of the stock dynamics. Consideration should be given to modelling M by size using the Lorenzen relationship and scaled to a mean value given by the Hamel method. This would avoid the need to model M by gender and would capture some of its annual variation.

vii. Thought needs to be given to the appropriate level of model complexity to ensure that the final base model fitted to the data also has the appropriate forecasting properties. A procedure needs to be developed to identify the most parsimonious model using an information statistic and the parameter correlation matrix.

viii. Although Stock Synthesis software (SS3) provides an impressive range of diagnostics to aid model development, in its present implementation it does not appear to provide the posterior distributions of the estimated parameters. This is something of a limitation as it hinders identifying problematic model fits and understanding the relative contribution of priors and data to the estimates. SS3 should be updated to provide full parameter posterior distributions.

ix. During the meeting a presentation was made that discussed observations suggesting the abundance of BDR in Oregon is of a similar order of magnitude to the black rockfish. This contrasts with the assessments for the two stocks which estimate the black rockfish to be considerably more abundant. Analyses carried out during the review did not provide sufficient support to change the base model but it remains an area for further research.

x. The assessments of BDR are split by state. Maps showing the distribution of blue and deacon rockfish suggest that blue rockfish predominate in the south and deacon in the north. The change in relative proportions seems to occur south of San Francisco. While the biology of the species may appear similar, there may be advantages in performing assessments north and south of the San Francisco area, rather than by state, to try to capture a more homogeneous species composition for each assessment.

xi. The review meeting was constructive and productive with effective chairing and excellent co-operation from the STAT teams. Meeting facilities were good and the local staff provided great support to the reviewers. There were no major disagreements between Panel members or the STATs.

Background

The National Marine Fisheries Service and the Pacific Fishery Management Council held a stock assessment review (STAR) panel meeting in July 2017 to evaluate and review benchmark assessments of Pacific coast blue/deacon rockfish and scorpionfish stocks.

Fish that were previously identified as blue rockfish have recently been found to consist of two species: blue and deacon rockfish. However, as it is not possible to separate the historical 'blue rockfish' landings and they have similar biology, these two stocks were assessed as one. Blue/deacon rockfish (BDR) are highly-valued by recreational fishermen, and rank among the five most important recreationally-caught groundfish in both Oregon and California. Blue rockfish (including deacon) was last assessed in 2007.

California scorpionfish is an important groundfish species for near-shore commercial and recreational fleets in southern California, as well as non-extractive uses such as *in situ* viewing (e.g. diving). Total catches have reached near the OFL over the past few years. The stock was last assessed in 2005 using Stock Synthesis 2, and OFLs/ACLs have been set based on a constant catch until a new assessment can be conducted.

The technical review of pre-STAR assessments took place during a formal, public meeting of fishery stock assessment experts from 24th-28th July. Two CIE reviewers were included in the review panel. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2.**

Description of the Individual Reviewer's Role in the Review Activities

Materials for the review were made available on the 11th July. These were studied prior to the meeting in preparation for the review. During the meeting, the reviewer took an active role in discussions and acted as rapporteur for the California BDR assessment. Requests for additional analyses for the STAT were noted and responses collated into a summary for the STAR panel report. The summary was prepared and sent the panel chair on the 2nd August. Comments on the draft STAR Panel report were sent to the Panel chair on the 9th August.

Summary of Findings for each ToR

Become familiar with the draft stock assessment documents, data inputs, and analytical models along with other pertinent information (e.g. previous assessments and STAR panel report when available) prior to review panel meeting.

Three draft stock assessment documents were reviewed. These covered assessments of BDR in California and Oregon, and Scorpionfish. In addition, material relating to the separation of blue and deacon rockfish, ROC analysis, M priors and previous STAR panel reviews was studied.

Discuss the technical merits and deficiencies of the input data and analytical methods during the open review panel meeting.

All the assessments use data quantifying total catches by fleet, indices of abundance, length compositions and age compositions.

Catch data

These data are perhaps the most important input to the assessment as they provide information on fishing mortality and help scale the assessment to the real fishable biomass. Each assessment attempts to characterize removals dating back to the beginning of the fishery which is at least the early 20th century. It is generally considered that "good" catch data are available from the 1980s onwards and that the early data are subject to much uncertainty. This is in part the result of the way species were recorded historically where species specific identification was absent, and due to the problems in quantifying removals from the recreational fleet. A considerable amount of effort has gone into the reconstruction of the catch time series, but ultimately it is reliant on pragmatic assumptions about the development of the fishery and associations between species of rockfish in official records. The recreational catch data are estimated by surveys of participating vessels and the MRFSS program (which forms one of these) has been criticized for its design. These uncertainties mean that the catch data are subject to error and probably bias, a problem that perhaps deserves greater consideration in the assessments. These, for example, assume the catch is fixed (and by implication error free) which means errors and bias in the data are forced into the estimates of stock biomass and exploitation rate. While it is not possible to recover accurate data from historical records, a study which attempted to quantify likely uncertainty and bias would make a valuable contribution to understanding the veracity of the estimated stock trends from the assessments and provide a basis for well-designed sensitivity tests.

Fishery dependent CPUE abundance data

While there are some fishery independent survey data, much of these data are opportunistic or not designed for the species assessed here. The STATs have therefore resorted to commercial CPUE data to construct indices of abundance to inform the assessments. These have been derived from various components of the recreational fleets through official records or observer programs. The main challenge in constructing such indices is in the removal of bias resulting from the way vessels target fish and to account for effort that is relevant to the species of interest in the assessments. All three assessments adopted a similar approach by identifying trips that might be expected to encounter the species concerned. An important element of this was to find the fish species in trips that were associated with the assessed species and then apply a filter that selected these trips. The STATs used a method proposed by Stephens and MacCall (2004) which uses ROC analysis to select a threshold for filtering the data based on the probability that the species of interest will occur in the catch given the species composition of the trip. Importantly, this process should help quantify the occurrence of zero catches when a trip was in an area likely to catch the target species. After filtering, a linear modelling approach was used to derive a CPUE index using either a delta-lognormal or negative binomial model to extract an abundance signal. These models considered effects such as year, area and season with the "best" model being chosen on the basis of the AIC. CVs on the index derived from the best model were then calculated separately in a Bayesian modelling package. There may be some redundancy in refitting the model rather than simply using the Bayesian approach to estimate the model and the CVs, but there are practical computational reasons for adopting the approach used by the STATs and any inconsistency should be minor.

Overall, the approach to estimating CPUE abundance indices was thorough and appears to be the best available given the data. Disappointingly, the resulting indices did not appear to contain much population signal and tended not to contribute much to the estimated stock biomass trajectory. In the

Oregon BDR assessment, for example, removing all the surveys from the assessment made no difference to the estimated biomass indicating that the catches drive the assessment.

Fishery independent abundance indices

For scorpionfish a number of surveys were available. These include a state trawl survey (SCBS), a university gillnet survey, public works (POTW) trawl survey and a NMFS trawl survey. None are specifically designed for scorpionfish, but provide potentially useful information. The POTW is the longest time series, but comprises a series of limited areas which may not adequately cover the population. The SCBS provides data every 4-5 years while the gillnet survey does not provide an index beyond 2008. The indices for these surveys are based on the delta lognormal model. The NWFSC trawl survey is likely to miss the inshore areas. The index is derived from a geostatistical model. The methods used to derive the indices are all well established and had been applied appropriately given the limitations of the survey designs. Taken together these surveys are important sources of fishery independent data.

A juvenile survey index is available for the California BDR. There is no strong year class signal in the index, though the model estimates are consistent with the relatively low 2005 year class and the stronger 2013 year class.

Length compositions

While some data from early years are available, the bulk of the length composition data began in the 1980s. Annual sample sizes at fleet level are generally small and reach a maximum of 3000-6000 fish per year in the most sampled fleet. Length compositions provide one of the few sources of data that can inform the model about year class strength and can be influential in estimating recruitment deviations in the model. Given that most data are for the post 1980s period, it means that there is very little information on age structure for the early period of the assessment. With the uncertainty in the early catch data, the reliability of the estimated stock trends pre-1980 is open to question.

Age compositions

A very limited amount of age data is available for all three assessments. For California BDR data are available from opportunistic sampling in 2010-2011 and from the cooperative survey from 1980-1984. Sample sizes are very small and amount to around 300 fish per year. For Oregon BDR samples, mostly from 2008-2015 are available with similar sample sizes. In the case of scorpionfish, age data from 2005-2016 are available, but again the sample sizes are small, amounting to less than 100 fish each year. Since the assessment model is age structured, age data is important in providing the model with information on recruitment deviations. Age structured data is most effective when a year class is sampled regularly throughout its life time so that an accurate picture of its survival rate can be estimated. This tends to be lacking in the data for these assessments.

Evaluate model assumptions, estimates, and major sources of uncertainty.

Model framework

All three assessments make use of the latest version of Stock Synthesis (SS3). This is a flexible modelling framework that can make use of a variety of disparate data and is particularly useful when time series data are discontinuous or where there are intermittent observations on length or age. It is therefore an appropriate choice for the assessments considered at the meeting.

Maximum likelihood forms the basis for parameter estimation but can be modified through the use of penalty functions that SS3 refers to as priors, though these are not applied in the true sense of Bayesian priors. The model is therefore founded in maximum likelihood, but leans toward a Bayesian approach by incorporating prior information. However, as currently implemented, parameter estimates are characterized by point estimates with approximate asymptotic variances rather than their posterior distributions. While MCMC sampling is available in SS3, analysts at the meeting indicated this was not fully functional. This is an important limitation especially where model parameters are not well estimated. There were some analyses which suggested that posterior parameter distributions may be multimodal and full posterior distributions would be an important diagnostic. Clearly, if distributions are not unimodal, then the interpretation of the model fit is problematic. In addition, where priors are applied, especially on parameters such as natural mortality or steepness, which are not well informed by data, comparing the posterior distribution to the prior is a useful tool in understanding information in the data.

Size composition model

The underlying population model is fully age structured, but it also models the size composition of the population. This is done by assuming growth follows a von Bertalanffy curve with dispersion around the mean. The size composition of the population is then reconstructed from the age composition using the length at age distribution. In the assessments considered here, observed length distributions were assumed to represent mid-year distributions with invariant growth rates. This inevitably raises the question as to whether this somewhat rough growth assumption is sufficiently robust in the light of real changes in growth by cohort, month and year. Fits to the length compositions were often poor and may reflect over-simplified modelling assumptions or poorly sampled length distributions. As length compositions are likely to be influential in the estimation of recruitment deviations (especially where age data are few or absent), this is an issue that merits further investigation.

Model parsimony

Each stock assessed reported the parameters that were estimated. These were generally in the region of 100 though they omitted survey catchability, q, which while estimated in closed form, nevertheless are model parameters and need to be considered. The number of parameters is large when considering the available data and there are clearly correlations between them. One would expect, for example correlations between R0, q and M which may be very high, and would indicate redundancy. Effort to try to find the most parsimonious model might help in reducing complexity and in identifying a model that had better predictive properties. There seemed to be a danger that while the chosen base models best fit the data, they may not have good forecasting ability given the high level of parameterization. It would be useful to compute an information statistic such as AIC to try to identify the best model, though this is made problematic by the use of penalty functions in the likelihood which affects the effective number of parameters to be estimated.

Selectivity

An important element of the SS3 approach is the need to model selectivity. The selectivity curves filter the length composition of the underlying population to explain the observed fleet specific length compositions. Selectivity is likely to change over time as management measures are applied. The approach adopted for these assessments was to use time blocks for fleets, where such regulations are thought to have affected selectivity. Clearly, it is desirable to model changes in selectivity to avoid miss-specification but this comes at a cost of increasing the number of parameters to be estimated. In addition, the choice of blocks may not capture the response of fleets to a variety of different factors that may go beyond management measures alone. An alternative is to allow the selection pattern to evolve over time using an auto-correlated random effect to smooth the data rather than force a rigid parametric form.

The shape of the selection curve at older ages or larger lengths can be significant in determining the scale of the estimated biomass. Inevitably these ages and lengths tend to be poorly sampled because they are less abundant. As a result, in fitting a selection curve, occasional observations at the largest size/oldest age may have undue influence. Choosing accumulator length bins or plus groups requires some thought. In the Oregon BDR assessment, for example, reducing the accumulator length bin and plus group gave a better fit to the data and is perhaps indicative for the need to choose these with care.

Natural Mortality, M

Natural mortality is included in the models either as a predetermined quantity or informed by a prior based on Hamel (2015). This, in effect, provides an estimate of the average annual non-fishing mortality experienced by an individual over its lifetime. M is generally size dependent (Lorenzen, 1996) while the models applied in these assessments assume it is fixed (except for gender differences). Using the Lorenzen relationship would imply, for example, that M for BDR Oregon would double over the size range 20-40 cm, which is the range accounting for most of the fish in the length composition data. Hence estimating M as a size/age invariant quantity will result in bias in the estimate of other model parameters and could be significant in the estimation of recruitment deviations. A possible way to address this issue is to use the Lorenzen relationship to characterize M by size and then scale the relationship to an overall mean given by the Hamel relationship.

The Hamel M value is calculated using an estimate of the oldest age and there was discussion at the meeting about what the best value to use was given the scarcity of observations at old ages and the problem of aging error which tends to be greater at high ages. In some cases the chosen oldest age for M estimation was lower than the oldest reported age in the samples. It seemed to me that while choosing lower age for the reasons of uncertainty mentioned earlier had some justification, it was nevertheless *ad hoc*. One approach might be to identify the highest age that there is confidence fish attain and then take an average of this and older observed ages using the estimated aging error to downweight older ages. In the case of BDR Oregon, the oldest estimated age for males was 29 while the age used for M was 26. This gives values of 0.21 and 0.19 respectively. Such a difference may appear trivial, but the male offset from female M in the model is only 0.05, so this will be heavily influenced by the choice of oldest age. Since the difference in M between males and females can be explained by size as well as maximum age, it might be simpler to model gender differences in M using a size relationship as described above.

The ability to estimate M within the model will depend on contrast in the data and constraints or assumptions on other parameters. It is usually difficult to estimate within the model because it is confounded with other parameters such as catchability. In view of the necessary simplifying assumptions on M, it is worth reflecting on whether trying to estimate its value is very useful. Inspection of the likelihood profiles over M for these stocks did not suggest that the various data sets provided consistent information, and in some cases implied multiple minima. In the case of California BDR, for example, the length data imply an M about half that of the age data, while the negative log-likelihood component of the index data reaches a maximum between these values.

A very useful sensitivity analysis to M was conducted for all three stocks and this perhaps is the most informative insight into the interpretation of stock status. For both BDR assessments, estimated stock status can be reversed if M is lower than the base assessment. In scorpionfish this is less pronounced. For this species, the value of M is much higher (0.26) than BDR yet ecological evidence does not suggest high predation. The Lorenzen equation would suggest a lower value (0.16 for a 30cm fish) so it is possible the assessment may be over-optimistic about the level of depletion since these possibly low values of M were not explored.

It is worth remembering that for these stocks the estimated fishing mortality rate is much lower than M, so much of the stock dynamics will be driven by factors external to the fishery. Whatever the true level of M, it is likely to vary over time, and since M cannot be included in the model dynamically (as there are not data to support it) the interpretation of stock trends is extremely difficult. Some of the annual variation in average M could be captured using a size dependent relationship as described earlier, as this would, for example, result in higher values of mean M when a large year class enters the stock and lower values when the age composition contains a high proportion of older fish.

Weighting multinomial data

Length and conditional age compositions are modelled as multinomial distributions where sample size is a critical weighting factor in the likelihood. The problem of identifying the correct effective sample size is well known. It will be most pronounced when the actual number of samples is small because the variability in the observations will be greatest. In all three assessments sensitivity to the choice of weighting was investigated using Francis and harmonic mean weighting, and shown to be a significant source of uncertainty. In the case of California BDR changing weights is sufficient to change the perceived status of the stocks relative to the management reference points. Perhaps the lesson here is the need to increase the number of samples both to provide the assessment with better data and to reduce the sensitivity to choice of weights.

Beverton-Holt steepness

All three assessment models use the Beverton-Holt stock-recruitment function parameterized in terms of steepness, h, and virgin biomass. For BDR in California, the estimated recruitment deviations suggest that the data have some information on steepness and when the model was run to estimate h, the resulting value was close to the fixed value of 0.718 used in the initial base run. However, the Oregon stock recruitment plot shows very little evidence of reduced recruitment at lower spawning biomass and scorpionfish, if anything, appear to show recruitment falling on the descending limb of a stock recruitment relationship. For these two stocks, steepness was fixed and probably represents the only sensible choice. It was suggested that scorpionfish recruitment may be correlated with warmer sea temperatures and an analysis performed during the meeting appeared to show a weak relationship between the estimated recruitment deviations and sea surface temperature. Such relationships are notoriously unreliable unless a convincing causal mechanism can be identified. Thus, while it may be worth pursuing the idea as a research topic, it is unwise to attempt to include such a relationship in the assessment model.

Sensitivity testing

Uncertainty was explored in a number of sensitivity tests. Reference has already been made to sensitivity to M and the choice of weights for the multinomial data. In addition, base models were run leaving out one dataset at a time, changing assumptions on R0 and steepness, h, and retrospective analysis. Typically, these analyses tend to show changes of scale rather than changes of overall trend or shape, though importantly in the case of California BDR, most sensitivity runs showed marked differences in the stock trend since the mid-1990s with the apparent strong stock recovery completely absent in some runs. It is also noteworthy that removing the Schmidt age data resulted in the model failing to converge. For this stock it appears that the early stock trend is insensitive to many of the likely uncertainties, but it is very difficult to establish the true stock status and trajectory in the most recent decade with any confidence. For Oregon BDR and scorpionfish there was also sensitivity to age data, illustrating its importance in the assessments. In these two stocks, the state of current depletion shows a wide spread as critical parameters are changed, but the year to year biomass changes are similar.

As noted earlier the catch data in all assessments are assumed to be exact or estimated with high precision. This is likely to be a necessary assumption for model convergence though it is clearly unrealistic. A sensitivity test that examined this assumption would be highly desirable. For Oregon BDR sensitivity to doubling or halving the historical catch was investigated. However, this simply changes the scale and the more important question is whether annual changes in the catch are well estimated since these may alter the perceived stock trend. Random draws of possible catch streams from likely ranges of uncertainty would be a more demanding sensitivity test. Preserving the annual autocorrelation would be necessary and could be achieved by adding a random error to the nominal values.

Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified.

The preceding section discusses a number of the main issues relating to uncertainty. Below are a number of additional issues that arose during the review.

Stephens-MacCall filtering

The construction of CPUE indices involved the use of the Stephens McCall method to identify trips related to the target species. This uses ROC analysis to identify a threshold for assigning trips where the number of false negatives and positives is equal. The Panel was concerned that the resultant indices may be sensitive to the choice of threshold and the STATs were asked to investigate alternative thresholds where false negatives were half or double the false positives. These changes affected the scale of the calculated indices, but had very little effect on the trends showing that the choice of threshold is not critical.

Discard fleets

In all three stocks, the level of discarding is considered to be fairly low. Nevertheless, for completeness, the STATs had included estimates of discards including size compositions. In the model these were treated as a separate fleet. While a convenient way of handling the data, it breaks the linkage between the landings component of the catch and it would be preferable to model the fleet total catch and then apply a post catch filter to derive discards. Attempts were made at the meeting to address this change but there were technical difficulties in fully implementing it. Runs were made where the discard biomass was simply added to the fleet landings. These changes made little difference to the overall model runs. Since the size the of the discard catch is small in relation to landings this is not a major issue, but it is probably preferable to avoid modelling a separate fleet, both because it is unrealistic and because it adds additional parameters to the model when precise data are scarce.

Oregon BDR relative abundance

During the meeting a presentation was made that discussed observations suggesting the abundance of BDR in Oregon is of a similar order of magnitude to the black rockfish. This contrasts with the assessments for the two stocks which estimate the black rockfish to be considerably more abundant than BDR. A series of additional runs were performed where R0 from the 2015 black rockfish assessment was used as a prior for R0 in the BDR assessment with a range of CVs. This showed that with an informative prior, the BDR assessment could be rescaled to the black rockfish level. The total log-likelihood increased by about 2.5 units, but as the CV prior on R0 increased, the estimated biomass reverted to a similar level to the base model. This suggests there is some weak information on scale in the BDR assessment pointing to a lower relative biomass compared to black rockfish. A further analysis looked at data from the onboard observer program to compare CPUE of BDR and black rockfish in similar habitat. This also suggested much lower relative abundance of BDR. Together these analyses did not provide sufficient support to change the base model but it remains an area for further research.

California BDR CPFV index

It was noted that the calculation of the CPFV index may be affected by the implementation of MPAs in more recent years. The STAT team recalculated the index using only areas outside the MPA, which tend to have lower catch rates. The new index had a lower scale but the time trajectory was very similar. It was agreed that this change would be included in a revised base model, however.

Scorpionfish fleet selectivities

The impingement survey was assumed to have constant selectivity in the pre-STAR base model and fit to the length compositions were poor. This was improved by estimating selectivity with a normal curve. This improved the fit to the data and was adopted for a revised base run. It should be noted, however, that this increases the number of parameters in the model and that the assessment shows high sensitivity to this data set.

For the recreational dead discard fleet, the pre-STAR base model included a single time block for selectivity. The Panel requested a run with three time blocks to allow for management changes. Not surprisingly this improved model fit and was adopted for a revised base model. However, it is questionable whether discards should be modelled as a separate fleet, and whether the improved fit is justified by the increased number of parameters when the change to the estimated depletion is negligible. Adding time blocks is in effect increasing the number of fleets without increasing the amount of data, so one might expect the additional parameters to be less well determined. This issue reinforces the need to examine a goodness of fit criterion such as AIC in guiding the choice of best model.

Determine whether the science reviewed is considered to be the best scientific information available.

The principal limitation in these assessment is the available data. Catch data pre-1980 are regarded as uncertain and there is a shortage of age data. The absence of a good quality fishery independent survey is also a major weakness. With these limitations in mind, the analyses are of a very high standard, making use of state-of-the-art analytical methods. I would judge the science to be the best available.

Stock Synthesis is now a well-established modelling framework and is well suited to the type and quantity of data available for assessment. It is, however, very complex both in the form of the objective function and the multiplicity of configuration options which can obscure what it actually is doing. By their nature, stock assessment models are over-parameterized and SS3 is no exception. With relatively uninformative data as in these assessments, the model is not well anchored and a wide variety of possible interpretations of the data are possible. Thus, while the science is of a high standard, the results of the assessments are not robust. At this point in time much more could be gained by collecting more informative data.

When possible, provide specific suggestions for future improvements in any relevant aspects of data collection and treatment, modeling approaches and technical issues, differentiating between the short-term and longer-term time frame.

BDR stock identity

Recent research has revealed the blue rockfish are comprised two species, blue rockfish and deacon rockfish. In the analyses presented the species were combined but separated into California and Oregon assessments. This was done for practical rather than biological reasons. Maps showing the distribution of the two species suggest that blue rockfish predominate in the south and deacon in the north. The change in relative proportions seems to occur south of San Francisco. While the biology of the species may appear similar, there may be advantages in performing assessments north and south of the San Francisco area to try to capture a more homogeneous species composition for each assessment. The current assessment areas are likely to reflect largely deacon rockfish for Oregon but a mixture of both species in California.

Data

At present, there is a large investment of analytic effort into somewhat limited data. More resources devoted to data collection would be highly beneficial. Priority should be given to the collection and processing of more age samples from all three species. This needs to be maintained to create a long time series of observations. While heroic attempts have been made to derive abundance indices from fishery data, these do not appear to be very informative and a dedicated survey would help overcome this problem. If possible, a fishery independent survey should be developed to calibrate estimates of biomass in the most recent year. The need for this is very evident in the California BDR assessment, where recent stock trends are not well determined and are sensitive to the inclusion/exclusion of data. During the meeting, it was suggested acoustic methods to assess biomass may be possible for BDR and there may be value in following up to see if this is indeed feasible.

Modelling approach

The use of SS3 allows highly complex and parameter rich models to be developed and the assessment models used in the assessments reviewed fall into this category. During the meeting there was a tendency to increase model complexity beyond the base models by estimating more selectivity parameters without increasing the number of observations. Not surprisingly, the fit to the data improves but this does not necessarily mean a better model. In general, while exploring complex models is undoubtedly useful, there should a systematic attempt to reduce complexity by critically examining the precision and posterior distributions of the parameters as well as their correlations. This would help in identifying redundancy and may help in improving model stability and predicative power. Time blocking of selectivity curves may help reduce residuals, but the question of whether the parameters of the curves were significantly different needs to be explored. A non-parametric time series approach may be a better way to capture time varying selection without over-parameterizing the model.

The assessment models chosen are likely to be the best way of incorporating a variety of different data into a comprehensive analysis. However, there may be some merit in applying other simpler approaches to gain insights into the information content of the data and identify conflicting signals. One might, for example attempt a stage based model where the population is modelled as two (or three) size components rather than a full length frequency (e.g. Mesnil, 2003). This would simplify assumptions on growth and selectivity. The CatchMSY method suggested by Martell and Froese (2013) may provide a very simple way of identifying plausible stock trajectories that are consistent with the catch stream and life history traits.

Preceding discussion considered various aspects of the way M is handled in the model. The current approach assumes constant M by gender based maximum observed age. In some model runs, M was estimated based on a prior. Since M is likely to be size dependent, a hybrid approach to modelling M may be more realistic by combining the size dependent relationship reported by Lorenzen (1996) with the scale proposed in the Hamel (2015) method. This might avoid the need to estimate separate M by gender, reduce bias in the estimation of selectivity parameters and capture some of the annual variability in average M associated with changing age compositions. Estimating M as a size independent constant within the model is probably unnecessary and simply provides more flexibility to an already highly parameterized model. Deriving M externally and then performing sensitivity analysis over a plausible range of M may be more useful.

Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

The review was conducted in a constructive manner and the STAT teams were responsive to the requests from the Panel for additional analyses with all the essential runs being completed during the meeting.

Many of the issues discussed have been referred to in earlier sections of this report. These included:

- Identifying selectivity assumptions that better explained the observed data
- Verifying that abundance indices were insensitive to the Stephens-MacCall filtering procedure
- Modelling discards as a separate fleet
- Evaluating the abundance of Oregon BDR relative to black rockfish
- Appropriate assumptions on natural mortality, particularly the best approach to using the Hamel method for deriving M
- Checking the influence of MPAs in the CFPV abundance index

Towards the end of the meeting, there were discussions on the states of nature for decision tables.

Overall there was effective engagement from all members of the Panel, the STATs and the Panel advisors. This lead to improvements in the configuration of the base models.

Recommendations for future assessments are discussed in the next section.

Conclusions and Recommendations

The assessment of BDR and scorpionfish represent the best science available given the existing data. The analyses were thorough and considerable work had gone into making good use of data from a variety of sources. The limited amount of age data and lack of informative abundance indices means that despite the elegance of the assessments, it is difficult to have high confidence in the estimated stock status. Not only are the estimated confidence intervals wide, but sensitivity analyses indicate that alternative explanations of the data are possible, which change perceived status. The problem is most severe in the case of California BDR, where the estimated stock recovery is highly uncertain.

Should managers attach importance to these stocks, then I would **recommend that priority be given to the collection and processing of more age samples from all three species. This needs to be maintained to create a coherent time series of observations. If possible a fishery independent survey should be developed to calibrate estimates of biomass in the most recent year.**

In common with many other assessments in this region, early catch estimates are subject to considerable uncertainty. The assumption that catches are exact and treated as fixed in the model is probably necessary but clearly unrealistic. Sensitivity to this problem needs to be adequately investigated as the catch data are influential in the assessment. I **recommend that a demanding sensitivity analysis is performed where plausible alternative catch streams are generated stochastically**. The current practice of halving or doubling the catch as a sensitivity test is not very demanding and is unlikely to probe the nature of the uncertainty in the data.

Natural mortality is the largest component of total mortality in these stocks and will drive much of the stock dynamics. I was not entirely convinced that modelling M as a constant value by gender was the best approach, or that trying to estimate its value within the model was useful. I **recommend that the way M is modelled and estimated is reviewed. Consideration should be given to modelling M by size using the Lorenzen relationship and scaled to a mean value given by the Hamel method.** This would avoid the need to model M by gender and would capture some of its annual variation.

I recognize that SS3 is a powerful, useful and appropriate tool for the assessment of these stocks. However, thought needs to be given to the appropriate level of model complexity to ensure that the final base model fitted to the data also has the appropriate forecasting properties. I would **recommend that a procedure is developed to identify the most parsimonious model using an information statistic and the parameter correlation matrix.**

SS3 provides an impressive range of diagnostics to aid model development. In its present implementation, it does not appear to provide the posterior distributions of the estimated parameters. This is something of a limitation as it hinders identifying problematic model fits and understanding the relative contribution of priors and data to the estimates. I **recommend that SS3 be updated to provide full parameter posterior distributions.**

Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

Draft assessment documents and supporting material were made available on the Pacific Fisheries Management Council ftp site two weeks in advance of the meeting. This is realistically the minimum advance time to review the assessments adequately. The principal documents are voluminous and take time to digest. As always, more time would be appreciated and would lead to more considered interventions at the review meeting. Understandably, there is a compromise to be struck between the completion of assessment documents and time available for review. Perhaps the two week period is the best that can be achieved.

The meeting itself was constructive and productive with effective chairing and excellent co-operation from the STAT teams. Meeting facilities were good and the local staff provided great support to the reviewers.

There is, perhaps, some blurring of the distinction between "review" and "analysis" in the way the assessments are developed during the STAR Panel meeting. The updating of base models with input from reviewers could be seen as the reviewers contributing to the assessment, in which case who reviews the new base model? It is, of course, appropriate for reviewers to make constructive suggestions, but the adoption/rejection of these needs to remain in the hands of the STAT, and it should be made clear the reviewers are not there to "agree" an assessment.

References

Hamel, O. 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. ICES Journal of Marine Science 72: 62-69.

Lorenzen, K. (1996). The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology, 49, 627–647.

Martell, S. and Froese, R. (2013) A simple method for estimating MSY from catch and resilience. Fish and Fisheries 14, 504–514.

Mesnil, B. (2003). The Catch-Survey Analysis (CSA) method of fish stock assessment: an evaluation using simulated data. Fisheries Research 63, 193–212.

Stephens, A., and MacCall, A. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research 70: 299-310.

Annex 1: Bibliography of materials provided for review

The following materials were made available in the PFMC ftp site before and during the meeting. They can be found at: ftp://ftp.pcouncil.org/pub/GF_STAR3_2017_Blue_Deacon_CAScorp/

Background

Design Considerations Dorn.pptx
STAR3 Specific-20170711T154402Z-001.zip
Shelton2014glmm.pdf
Steepness_prior_for_SSC_2017_V4.pdf
Summary Comparison of TriennialShelf and Shelf_Slope Surveys.doc
Thorson and Lewis_ Comparing estimates of abundance trends.pdf

Blue/Deacon Rockfish (BDR)

BDR_OR_SSfiles.zip
Blue-Deacon 2017 Pre-STAR Draft, 2017-07-11.pdf
Blue_Rockfish_Total_Mortality_Reports_2017-06-27.xlsx
OR_BDR_Acoustics_appendix.docx
OR_BDR_ROV_survey_appendix.DOCX
OR_BDR_Substrate appendix.docx
OR_BDR_hook_line_appendix.docx

Scorpionfish

California_scorpionfish_2017.pdf
Scorpionfish_preSTAR_basemodel_2017.zip
Scorpionfish_preSTAR_basemodel_2017.zip

Requests

1st Set of Requests to the BDR-OR STAT.docx
1st Set of Requests to the CA Scorp STAT.docx
2nd Set of Requests to the BDR-CA STAT.docx
2nd Set of Requests to the BDR-OR STAT.docx
2nd Set of Requests to the CA Scorp STAT.docx
3rd Set of Requests to the BDR-CA STAT.docx
3rd Set of Requests to the BDR-OR STAT.docx
3rd Set of Requests to the CA Scorp STAT.docx

Presentations

BDR CA STAT responses to STAR Requests 7 (revised), 8, and 9.pptx

BLUE_DEACON_STAR_OR_.pptx
CA BDR STAT responses to STAR Requests 1-7.pptx
CAscorp_fleets.pdf
CDFW Aerial Survey Kelp Index 2002-2016_FINAL.docx
OR_bluedeaconcomments.doc
PostSTARBaseModel_BDR_OR.zip
Responses_STAR-Request1_OR_BDR.pptx
Responses_STAR-Request2_BDR_OR.pptx
SCOR_2017_STAR_presentation.pdf
STAR - Blue-Deacon CA Data 2017-07-24.pptx
STAR - Blue-Deacon CA Model 2017-07-24.pptx
STAR - Blue-Deacon Overview 2017-07-24.pptx
STAR-Data_OR_BDR.pptx
STAR-Model_OR_BDR.pptx
STAR-Request3_OR.pptx
STARBase3.zip
Scorpionfish Request STAT Responses.pptx (2).pdf
Scorpionfish Request STAT Responses_2.pptx.pdf
T_Thompson_Deacon Rock.docx

Annex 2: Statement of Work

External Independent Peer Review by the Center for Independent Experts

Stock Assessment Review (STAR) Panel 3

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf). Further information on the CIE program may be obtained from www.ciereviews.org.

Project Description:

The National Marine Fisheries Service and the Pacific Fishery Management Council will hold stock assessment review (STAR) panels in 2017 to evaluate and review benchmark assessments of Pacific coast groundfish stocks. The goals and objectives of the groundfish STAR process are to:

1) ensure that stock assessments represent the best available scientific information and facilitate the use of this information by the Council to adopt OFLs, ABCs, ACLs, (HGs), and ACTs;
2) meet the mandates of the Magnuson-Stevens Fisheries Conservation and Management Act (MSA) and other legal requirements;
3) follow a detailed calendar and fulfill explicit responsibilities for all participants to produce required reports and outcomes;
4) provide an independent external review of stock assessments;
5) increase understanding and acceptance of stock assessments and peer reviews by all members of the Council family;
6) identify research needed to improve assessments, reviews, and fishery management in the

future; and

7) use assessment and review resources effectively and efficiently.

Fish that were previously identified as blue rockfish have recently been determined to consist of two species: blue and deacon rockfish. Because there is no way to separate the historical 'blue rockfish' landings and they seem to have similar growth rates, these two stocks are being assessed as one. Blue/deacon rockfish are highly-valued by recreational fishermen, and rank among the 5 most important recreationally-caught groundfish in both Oregon and California. Blue rockfish (including deacon) was last assessed in 2007.

California scorpionfish is an important groundfish species for near-shore commercial and recreational fleets in southern California, as well as non-extractive uses such as *in situ* viewing (e.g. diving). Total catches have reached near the OFL over the past few years with the average percent attainment of the OFL (e.g. catch/OFL) of 95%. The stock was last assessed in 2005 using Stock Synthesis 2, and OFLs/ACLs have been set based on a constant catch until a new assessment can be conducted.

These assessments will provide the basis for the management of the blue/deacon rockfish and California scorpionfish stocks off the West Coast of the U.S., including providing the scientific basis for setting OFLs and ABCs as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day meeting of fishery stock assessment experts. Participation of external, independent reviewer is an essential part of the review process. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers

NMFS requires two CIE reviewers to participate in the stock assessment review panel. One CIE reviewer shall conduct an impartial and independent peer review of the assessments described above and in accordance with the SoW and ToRs herein. Additionally, a second "consistent" CIE reviewer will participate in all STAR panels held in 2017 and the SOW and ToRs for the "consistent" CIE reviewer are included in a separate SoW (See **Attachment A**).

Both CIE reviewers shall be active and engaged participants throughout panel discussions and able to voice concerns, suggestions, and improvements while respectfully interacting with other review panel members, advisors, and stock assessment technical teams. The CIE reviewers shall have excellent communication skills in addition to working knowledge and recent experience in fish population dynamics, with experience in the integrated analysis modeling approach, using age-and size-structured models, use of Markov Chain Monte Carlo (MCMC) to develop confidence intervals, and use of Generalized Linear Models in stock assessment models.

Statement of Tasks

The CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Pre-review Background Documents: At least two weeks before the peer review, the contractor will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Documents to be provided to the CIE reviewers prior to the STAR Panel meeting include:

- The current draft stock assessment reports;
- The Pacific Fishery Management Council's Scientific and Statistical Committee's Terms of Reference for Stock Assessments and STAR Panel Reviews;
- Stock Synthesis (SS) Documentation
- Additional supporting documents as available.
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer).

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein.

Contract Deliverables - Independent CIE Peer Review Reports: The CIE reviewers shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report: The CIE reviewers may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewer are not required to reach a consensus, and should provide a brief summary of each reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Timeline for CIE Reviewers

The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided in advance of the peer review.
2) Participate during the STAR Panel review meeting **scheduled in Santa Cruz, California during the dates of July 24-28, 2017** as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than August 11, 2017, the CIE reviewer shall submit their independent peer review report to the contractor. The CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each ToR in **Annex 2**.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

For the **STAR panel 3** review, each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in **Santa Cruz, California during the dates of July 24-28, 2017**.

Period of Performance

The period of performance shall be from the time of the award through September 15, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables

The contractor shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

June 19, 2017	Contractor selects and confirms reviewers
July 10, 2017	Contractor provides pre-review documents to the reviewers
July 24-28, 2017	Each reviewer participates and conducts an independent peer review during the panel review meeting
August 11, 2017	Contractor receives draft reports
August 22, 2017	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content in **Annex 1**; (2) The reports shall address each ToR as specified **Annex 2**; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $7,700.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contacts
Stacey Miller, NMFS Project Contact
National Marine Fisheries Service,
2032 SE OSU Drive
Newport, OR 97365
Phone: 541-867-0535

Jim Hastie
National Marine Fisheries Service,
2725 Montlake Blvd. E,
Seattle WA 98112
Jim.Hastie@noaa.gov
Phone: 206-860-3412

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Stock Assessment Review (STAR) Panel 3

1. Become familiar with the draft stock assessment documents, data inputs, and analytical models along with other pertinent information (e.g. previous assessments and STAR panel report when available) prior to review panel meeting.
2. Discuss the technical merits and deficiencies of the input data and analytical methods during the open review panel meeting.
3. Evaluate model assumptions, estimates, and major sources of uncertainty.
4. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified.
5. Determine whether the science reviewed is considered to be the best scientific information available.
6. When possible, provide specific suggestions for future improvements in any relevant aspects of data collection and treatment, modeling approaches and technical issues, differentiating between the short-term and longer-term time frame.
7. Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Annex 3: Tentative Agenda

TBD

Stock Assessment Review (STAR) Panel 3

NMFS Southwest Fisheries Science Center

110 Shaffer Road

Santa Cruz, California

July 24-28, 2017

Annex 3: Panel membership and participation

STAR Panel Members

Martin Dorn, National Marine Fisheries Service Alaska Fisheries Science Center, Scientific and Statistical Committee, STAR Panel Chair
Panayiota Apostolaki, Center for Independent Experts
Robin Cook, Center for Independent Experts
Owen Hamel, National Marine Fisheries Service Northwest Fisheries Science Center, Scientific and Statistical Committee

Stock Assessment Team (STAT) Members (BDR)

E.J. Dick, National Marine Fisheries Service Southwest Fisheries Science Center, California Blue/Deacon Rockfish STAT Lead
Aaron Berger, National Marine Fisheries Service Northwest Fisheries Science Center, Oregon Blue/Deacon Rockfish STAT Lead
Brett Rodomsky, Oregon Department of Fish and Wildlife

Stock Assessment Team (STAT) Members (Scorpionfish)

Melissa Monk, National Marine Fisheries Service Southwest Fisheries Science Center, STAT Lead
John Budrick, California Department of Fish and Wildlife, Scientific and Statistical Committee
Xi He, National Marine Fisheries Service Southwest Fisheries Science Center

STAR Panel Advisors

Patrick Mirick, Oregon Department of Fish and Wildlife, Groundfish Management Team
Louie Zimm, Groundfish Advisory Subpanel
John DeVore, Pacific Fishery Management Council